Week 13: Activity

Last Week

- Spatial EDA
- GP models to spatial data
- Spatial Prediction / Model Choice
- Anisotropic Spatial Models

This Week

- GLM models
- Spatial GLMs

Generalized Linear Model Notation

There are three components to a generalized linear model:

1. Sampling Distribution: such as Poisson or Binomial

2. Linear combination of predictors: $\eta = X\beta$

3. A link function to map the linear combination of predictors to the support of the sampling distribution.

Binary Regression Overview

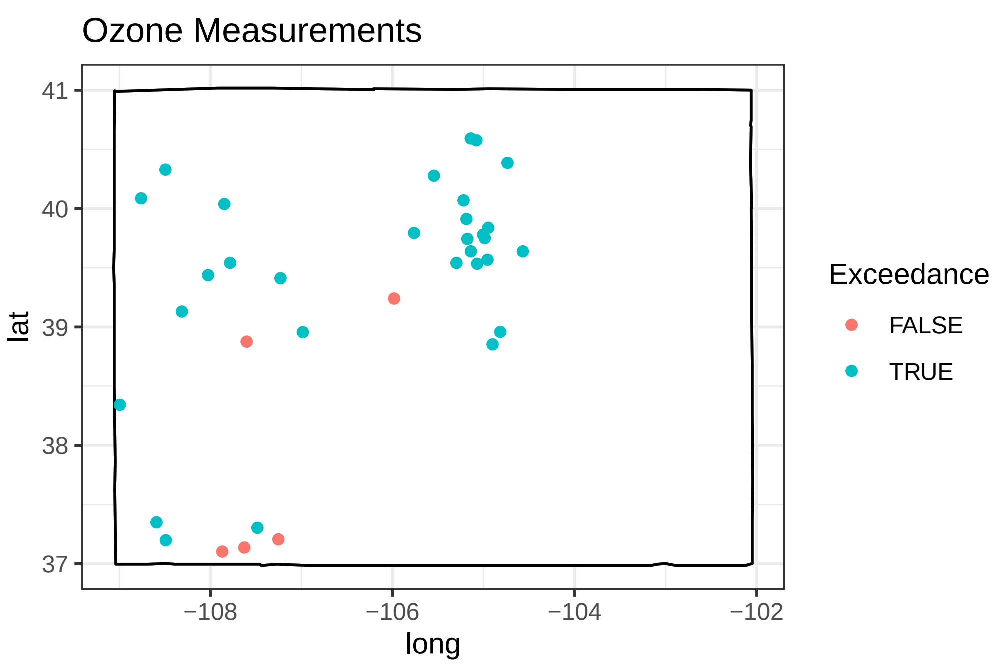
Write out the complete model specification for binary regression.

Latent interpretation of probit model:

Let $z_i > 0$ if $y_i = 1$. Otherwise, let $z_i < 0$. Then $z_i \sim N(X\beta, 1)$ is a latent continuous variable that is mapped to zero or 1.

For a set of predictors, X', then $z' \sim N(X'\beta, 1)$ and the probability of a 1 or zero can be obtained by integrating the latent distribution.

Consider air quality data from Colorado as a motivating example.



Interpret the output.

```r
CO <- CO %>% mutate(north = as.numeric(Latitude > 38 ))
glm(Exceedance~north, family=binomial(link = 'probit'),data=CO) %>% display()
```

```
glm(formula = Exceedance ~ north, family = binomial(link = "probit"),
    data = CO)
            coef.est coef.se
(Intercept) 0.00     0.51
north       1.48     0.62
---
  n = 35, k = 2
  residual deviance = 22.9, null deviance = 28.7 (difference = 5.8)
```

```r
glm(Exceedance~north, family=binomial(link = 'logit'),data=CO) %>% display()
```

```
glm(formula = Exceedance ~ north, family = binomial(link = "logit"),
    data = CO)
            coef.est coef.se
(Intercept) 0.00     0.82
north       2.60     1.10
---
  n = 35, k = 2
  residual deviance = 22.9, null deviance = 28.7 (difference = 5.8)
```

Spatial Binary Regression

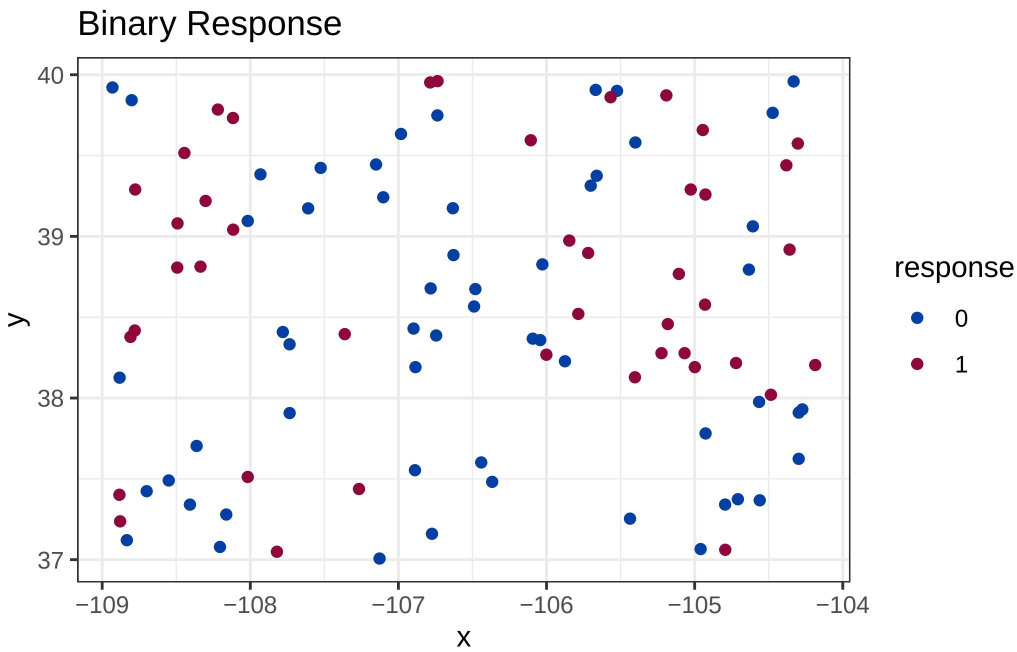
Assume $Y(s_i)$ is the binary response for s_i,

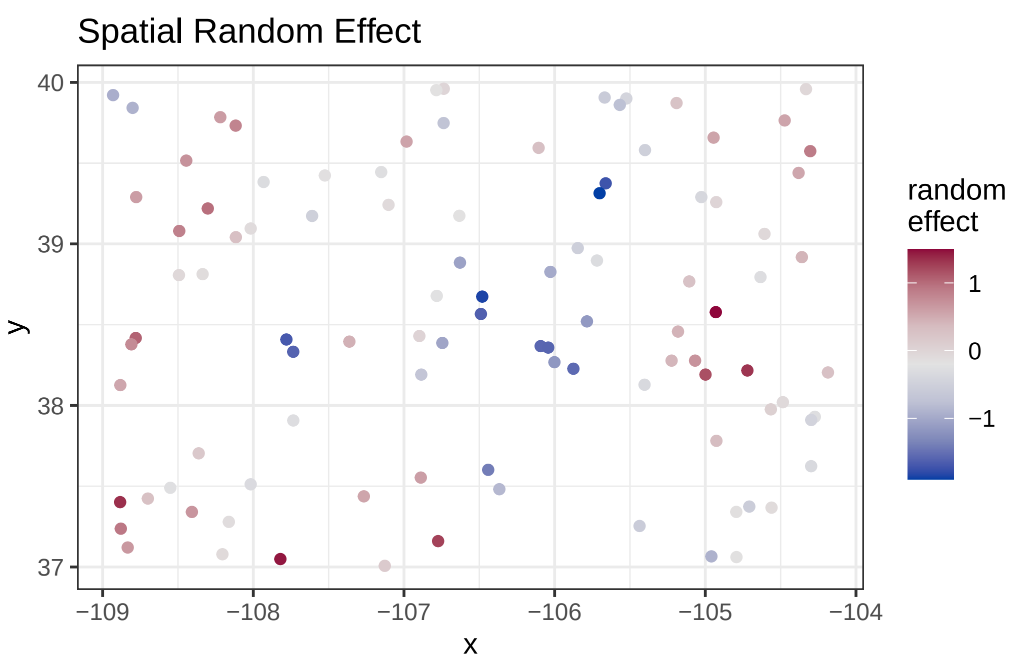
$$\begin{aligned} Y(s_i)|\beta, w(s_i) &\sim Bernoulli(\pi(s_i)) \\ \Phi^{-1}(\pi(s_i)) &= X(s_i)\beta + w(s_i), \end{aligned}$$

where $W \sim N(0, \sigma^2 H(\phi))$

Simulating spatial random effects for binary data

```r
N.sim <- 100
Lat.sim <- runif(N.sim,37,40)
Long.sim <- runif(N.sim,-109,-104)
phi.sim <- 1
sigmasq.sim <- 1
beta.sim <- c(-1,1)
north.sim <-  as.numeric(Lat.sim > 38)


d <- dist(cbind(Lat.sim,Long.sim), upper = T, diag = T) %>% as.matrix
H.sim <- sigmasq.sim * exp(- d / phi.sim)
w.sim <- rmnorm(1,0,H.sim)
xb.sim <- beta.sim[1] + beta.sim[2] * north.sim
y.sim <- rbinom(N.sim,1,pnorm(xb.sim + w.sim))
```


Binary Response
40
39
38
37
y
−109
−108
−107
−106
−105
−104
x
response
0
1


Spatial Random Effect
40
39
38
37
y
−109
−108
−107
−106
−105
−104
x
random
effect
1
0
−1

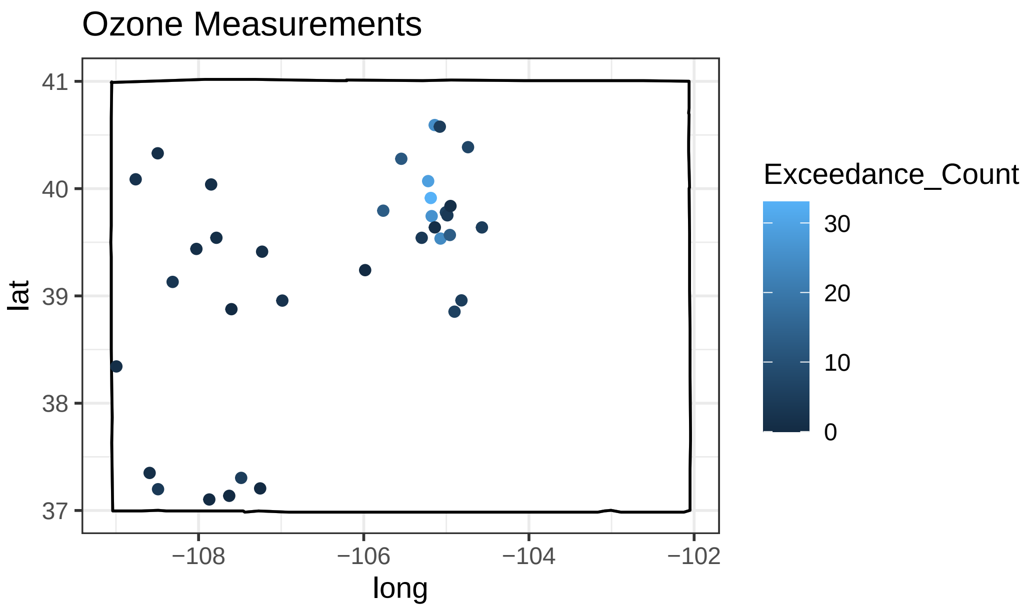
STAN: probit regression

```
// The input data is a vector 'y' of length 'N'.
data {
  int<lower=0> N;
  int<lower=0,upper=1> y[N];
  vector[N] x;
}

// The parameters accepted by the model. Our model
// accepts two parameters 'mu' and 'sigma'.
parameters {
  real beta0;
  real beta1;
}

transformed parameters {
  real<lower = 0, upper = 1> p[N];
  for (i in 1:N) {
    p[i] = Phi(beta0 + beta1 * x[i]);
  }
}


// The model to be estimated.
model {
  for (i in 1:N){
    y[i] ~ bernoulli(p[i]);
  }
}
```

Binary Regression

Spatial Poisson Regression

Motivation



Poisson Regression Overview

Write out the complete model specification for Poisson regression.

Assume Y_i is the count response for the i^{th} observation,

Next write out a Poisson regression model with spatial random effects

1. **Simulate and visualize spatial random effects for binary data: No Covariates**

2. **Fit a model for this setting**